We Claim:

1. A method of predicting a biological activity of a molecule, comprising:

obtaining spectral data for a test compound;

comparing the spectral data of the test compound to a pattern of spectral data associated with a biological activity, derived not exclusively from the assigned spectral data of a training set of compounds having a known biological activity;

detecting similarities between the pattern of spectral data associated with a biological activity of the training set and a pattern of spectral data for the test compound to determine whether the test compound is predicted to share the biological activity.

2. The method of claim 1 wherein the spectral data are obtained without first correlating the spectral data with corresponding structural features.

3 The method of claim 1 wherein the pattern of spectral data associated with a biological activity is derived without first correlating the spectral data with corresponding structural features.

4. The method of claim 1, wherein the pattern of spectral data of the training set is a pattern obtained by separating the spectral data of the training set of compounds into sub-spectral units.

5. The method of claim 4, wherein the pattern of spectral data of the test compound is obtained by separating the spectral data of the test compound into substantially the same sub-spectral units into which the spectral data of the training set is separated.

6. The method of claim 1, wherein the spectral data is one type of spectral data.

7. The method of claim 6, wherein the spectral data comprises one of nuclear magnetic resonance, mass spectral, infrared, ultraviolet-visible, fluorescence, or phosphorescence data.

8. The method of claim 1, wherein the spectral data is a composite of different types of spectral data.

9. The method of claim 8, wherein the different types of spectral data comprise two or more of the group consisting of nuclear magnetic spectroscopy (NMR), mass spectroscopy (MS), infrared (IR) spectroscopy, and ultraviolet-visible (UV-Vis) spectroscopy.

10. The method of claim 1, wherein the spectral pattern of the test compound and the spectral pattern of the training set are segmented into sub-spectral units, and the spectral data of the training set is scaled to normalize the importance of different signals within the spectral data of the training set prior to deriving a pattern associated with a biological activity.

11. The method of claim 10, wherein the scaling is auto-scaling.

12. The method of claim 10, wherein the spectral data of the training set is weighted to emphasize signals that are important for determining the endpoint class of compounds in the training set before deriving a pattern associated with a biological activity.

13. The method of claim 12 wherein the weighting is Fisher-weighting.

14. The method of claim 1, wherein detecting similarities between the pattern of spectral data associated with a biological activity of the training set and the pattern of spectral data for the test compound comprises performing computer implemented pattern recognition.

15. The method of claim 1, wherein detecting similarities between the pattern of spectral data associated with a biological activity of the training set and the pattern of spectral data for the test compound comprises detecting relative intensities of signals associated with one or more of the sub-units of the spectrum of the training set, and detecting relative intensities of signals associated with the same one or more sub-units of a spectrum of the test compound.

16. The method of claim 15, wherein the relative intensities are canonical variate factors of the spectral data associated with a biological activity of the training set and the spectral signals of the test compound.

17. The method of claim 1, wherein the method is computer implemented.

18. A computer implemented system for predicting a biological activity of a test compound, comprising:

receiving as input spectral data for a test compound;

receiving as input spectral data of a training set of compounds having a known biological activity; and

comparing the pattern of spectral data of the training set associated with the biological activity to the spectral pattern of the test compound to determine whether the spectral pattern of the test compound matches the spectral pattern associated with the biological activity of the training set.

19. The computer implemented system of claim 18, wherein comparing the spectral patterns comprises comparing the spectral patterns with computer implemented pattern recognition programs.

20. The computer implemented system of claim 19, wherein the spectral data for the test compound and the spectral data for the training set are divided into substantially identical spectral bins, so that a signal within individual spectral bins is compared between spectral patterns of the training set associated with the biological activity and the test compound.

21. The computer implemented system of claim 18, wherein the spectral patterns are obtained by inputting spectral data selected from the group consisting of nuclear magnetic resonance data, mass spectral data, infrared data, untraviolet-visible data, fluorescence data, phosphorescence data, and composites of two or more such spectral data.

22. The computer implemented system of claim 21, wherein the spectral data for the training set are converted into canonical variates associated with the

biological activity of the training set, and the spectral data for the test compound are compared to the canonical variates of the training set spectral data.

23. The computer implemented system of claim 22, wherein the biological activity is binding affinity to a hormone receptor, and the canonical variates for the training set include peaks in bins that are associated with hormone receptor binding of a pre-selected affinity.

24. The computer implemented system of claim 23, wherein the spectral data comprise nuclear magnetic resonance data and mass spectral data.

25. A computer readable medium having stored thereon instructions for performing the actions of claim 1.

26. A computer readable medium having stored thereon instructions for performing the actions of claim 18.

27. A method for predicting a biological, chemical, or physical property of a molecule, comprising:

providing spectral data segmented into spectral sub-units, for a plurality of training compounds;

inputting the segmented spectral data and endpoint data into a pattern-recognition program;

training the pattern-recognition program with the segmented spectral data and endpoint data to establish a relationship between the spectral sub-units of the segmented spectral data and the endpoint;

providing segmented spectral data for a test compound that is segmented into substantially the same spectral sub-units as the spectral data of the training compounds; and

comparing the relationship between the spectral subunits of the segmented spectral data and the endpoint to the spectral subunits of the test compound's segmented spectral data to predict the endpoint of the test compound.

28. The method of claim 27, wherein knowledge of the structures of the training compounds and the test compound are not necessarily known beforehand.

29. The method of claim 27 wherein the segmented spectral data of the training set is autoscaled and Fisher-weighted before training the pattern recognition program.

30. The method of claim 27 wherein the spectral data is chosen from the group consisting of nuclear magnetic resonance data, mass spectral data, infrared data, UV-Vis data, fluorescence data, phosphorescence data, and composites thereof.

31. The method of claim 30 wherein the spectral data is chosen from the group consisting of ^{13}C NMR data, EI MS data, and composites thereof.

32. The method of claim 27, wherein the endpoint is a ligand-target molecule-binding affinity.

33. The method of claim 32, wherein the ligand-target molecule binding affinity is an estrogen-receptor binding affinity.

34. The method of claim 27, wherein the endpoint is selected from the group consisting of:

a measure of biodegradability;

a measure of toxicity;

participation in a metabolic pathway;

a partition coefficient;

a reaction rate;

a quantum yield;

a measure of phototoxicity;

an equilibrium constant; and

a site of reaction on a molecular structure.

35. The method of claim 34, wherein the partition coefficient is the octanol-water partition coefficient.

36. The method of claim 27, wherein non-spectral structure descriptors that do not necessarily depend upon structural knowledge beforehand are provided for the test compound and the training compounds; used to establish a relationship between

the segmented spectral data, the non-spectral structure descriptors, and the endpoint for the training compounds; and used to predict the endpoint of the test compound.

37. The method of claim 36, wherein the non-spectral structure descriptors are chosen from the group consisting of partition coefficients, solubilities, relative acidities, relative basicities, pKa, pKb, reaction rates, and equilibrium constants.

38. The method of claim 37, wherein the partition coefficient is the octanol-water partition coefficient.

39. The method of claim 27, wherein non-spectral structure descriptors that do depend upon structural knowledge beforehand are provided for the test compound and the training compounds; used to establish a relationship between the segmented spectral data, the non-spectral structure descriptors, and the endpoint for the training compounds; and used to predict the endpoint of the test compound.

40. The method of claim 39, wherein the non-spectral descriptors are calculated using a quantum mechanical or electrostatic potential method.

41. A method for using spectral data as a set of structure descriptors for a compound that does not necessarily require knowledge of the compound's structure beforehand, comprising:

providing spectral data; and

segmenting the spectral data into bins.

42. A method for establishing a relationship between spectral data and a biological, chemical, or physical property, comprising:

providing spectral data;

segmenting the spectral data into bins.

detecting patterns in the bins of the spectral data that are associated with the property.

43. The method of claim 42, further comprising detecting corresponding patterns in spectral data of test compounds to select test compounds having the property.

44. The method of claim 43, wherein the test compounds are mixtures of compounds.

45. The method of claim 42 including one or more of: auto-scaling the segmented data; and weighting the segmented spectral data.

46. The method of claim 45 wherein weighting of the segmented data comprises Fisher-weighting of the segmented spectral data.

47. The method of claim 1, wherein the biological activity of the test compound is predicted without reference to a chemical structure of the test compound.

48. A method for establishing a spectral data activity relationship, comprising;

providing endpoint data for a plurality of compounds;

providing spectral data for a plurality of compounds;

segmenting the spectral data for the plurality of compounds into bins;

autoscaling the numerical data obtained from the spectral features within each of the bins;

Fisher-weighting the data within each of the bins; and

correlating information in the bins and the endpoint using a means for pattern recognition.

49. The method of claim 47 wherein the spectral data is selected from the group consisting of NMR data, MS data, IR data, UV-Vis data, fluorescence data, phosphorescence data, and composites thereof.

50. The method of claim 48 wherein two or more types of spectral data are normalized to each other in a composite.

51. A method for determining the structural features of a plurality of compounds that contribute to determining a particular endpoint property exhibited by the compounds, comprising:

providing segmented spectral data for the plurality of compounds;

providing endpoint data for the plurality of compounds;

establishing a spectral data-activity relationship by identifying the segmented spectral features that bias toward increased endpoint values and the segmented spectral features that bias toward decreased endpoint values; and

identifying the structural feature leading to the segmented spectral features that bias toward increased or decreased endpoint values for the plurality of compounds.

52. The method of claim 51 wherein the segmented spectral data is selected from the group consisting of NMR data, MS data, IR data, UV-Vis data, fluorescence data, phosphorescence data, and composites thereof.

53. The method of claim 52 wherein two or more types of segmented spectral data are normalized to each other in a composite.

54. The method of claim 52 wherein the segmented spectral data is a composite of ^{13}C NMR and EI MS data.